Deviation models of regulation: a knowledge-based approach

Ephraim Nissan

We propose a unified schema for representing, for the purposes of symbolic manipulation, several parallel legal corpora. The specificalities of each single corpus are represented as details deviating from the shared core. We resort to passive frames embodied in deeply nested relations, and to a meta-level augmented ruleset that specifies the structure of the nested relations. The approach can be interfaced with a logic representation at a lower level. Textual fragments given as value, in frames, can be parsed. The example discussed is based on Italy’s Statuti Regionali. While this legal application is still unimplemented, the elements of the architecture exposed have already been implemented in the framework of other projects I developed or directed.

1. THE APPLICATION

Italy has 20 regions; each region has its own statute, that has the role of a constitution for that particular region. Five regions have a special statute, and enjoy a higher degree of autonomy than the other 15 regions, which, in turn, have an ordinary statute, that was promulgated later. (Statuti Regionali 1972).

The most developed and recent — it is of 1963 — among the special statutes, is in force in Friuli-Venezia Giulia. Sicily enjoys the broadest autonomy. In Trentino-Alto Adige, the two provinces are each autonomous, because of the particular ethnic and linguistic situation. (ibid.: pp. 5-6). Sardinia and Valle d’Aosta, too, enjoy a special autonomy. As to ordinary statutes, they have been in force since 1972; they are often more detailed than the special statutes.

* Department of Mathematics & Computer Science, Ben Gurion University of the Negev, P.O. Box 653, Beer-Sheva 84105, Israel.
2. SEPARATE REPRESENTATION VS. A DEVIATION MODEL

Especially the ordinary statutes are rather similar. We are interested in finding out a kind of knowledge-representation schema, that would best suit them, and account for their relationship. Our suggestion is to provide a unified representation, with respect to which, deviations would be stated.

However, the decision whether to provide completely separate representations for every single statute, as opposed to the unified deviation model, should depend on the uses envisaged for the online tool. For comparative study, the deviation model would be suitable, but would already embody a comparative analysis. Instead, for applications concerned mainly with a given region, there is no need to have plurality represented per sé, and encoding each statute on its own would suffice.

Another choice has to be made, concerning the format of representation:
1) The text alone could be stored;
   – either it would be used just for traditional textual information retrieval,
   – or a powerful parser, capable of analyzing semantics beside syntax,
   and embodying also the basics of legal knowledge, should be assumed to
   be available for application on the original text of the statutes.

2) Otherwise – and this is the option we are concerned with – an internal semantic representation should be encoded manually. The original text would also be made available (for display purposes). A parser, too, would be provided, but just a simple one, able to analyze simple phrases or sentences with a restricted lexicon: this would simplify and enhance the appearance of the internal representation. The latter would have its data manipulated symbolically by artificial intelligence means, these being the internal workings of an expert system with a question-answering capability.

Related assumptions have to be made about the modalities of interaction. Alternatives follow.

1) Text could be scrolled sequentially, or searched through accesses performing pattern matching on strings or even semantic concepts (if a net of concepts is provided, along with synonymy associations); a concordance could be useful.

2) Interaction could be hypertextual: the user would be expected to select manually those associations whose availability is pinpointed on the display.

3) Interaction could be in natural or quasi-natural language, provided that a question-answering capability is made available, that includes the ability
- to analyze and synthesize natural language,
- to manipulate domain knowledge through a full-fledged inference and association capability in the application domain, and
- to explain responses.

3. EXAMPLES OF DEVIATION MODELS FROM OTHER AI PROJECTS

Before venturing on a technical explanation of the knowledge-representation schema we propose for the Italian regional statutes, let us consider two instances of application of a deviation model, in my own projects.

3.1. Deviation in the representation of lexical, morphological, and cultural knowledge of a family of natural languages

Knowledge-bases embodying common-sense knowledge about real-world kinds of objects or beings, or about abstract notions, usually fail to capture the perceptual-cultural diversity of the universe of concepts, and, when associating these concepts with terms, they fail (even in systems intended for machine-translation) to account for kinship between languages, or for the manipulations that both terms and notions have undergone. Not only: computer systems meant to process the lexicon of natural-language text or utterances, are not robust at all when faced with linguistic data they have incomplete knowledge about. There is no smooth deterioration of performance, and this is an often voiced criticism levelled at artificial intelligence applications in general. It is true, in particular, for systems purporting to embody some linguistic competence: human beings are often able to analyze and guess the meaning of a term they never met with before, and belonging to their own language; that ability depends on word-formation knowledge of the lexical and derivational morphemes available, in the natural language considered, for forming terms; besides, the semantic components that morphological analysis yields are matched, by the human subject, against a universe of concepts mapping his or her common-sense knowledge, until a sensible semantic concept is identified, that could be referred to by the new term. I have developed ONOMATURGE, an expert system that embodies a computational model of Hebrew word-formation, for the generation aspect: it proposes neologisms able to convey a given concept, and evaluates their psycho-linguistic clarity. I have described this expert system in (Nissan 1987 a,b). On the other hand, natural-language processing programs I am aware of, are unable to cope with terms they never met with before, unless
they interrogate the user about their meaning, as in the case of a system described by Haas and Hendrix (1984).

Polyglot knowledge, too, makes use of knowledge on the lexicon and the derivational morphology of different languages, to allow a reconstruction of meaning as conveyed by a foreign-language text (or speech), notwithstanding incomplete linguistic knowledge. This phenomenon is termed inter-intelligibility: knowing Italian, I can understand some Spanish, albeit I cannot generate correct Spanish sentences. The use of polyglot skills that rely on inter-intelligibility, often incurs in errors, because of false semantic correspondences between terms that look alike or as equivalent, and that are known as faux amis. Inter-intelligibility skills have not been modelled, so far, in artificial intelligence systems, albeit such skills make human receptiveness to linguistic messages usually more robust.

In machine-translation today, bilingual dictionaries (or, more efficiently, lexicons relating one internal representation to a fan of natural languages) tend to state lexical equivalences without exploiting word-formation knowledge that would allow some inter-intelligibility; this way, only those equivalences that are stated in the lexicon can be exploited for the purposes of translation (the whole of potential performance at establishing lexical correspondences is conceivable as the sum of the correspondences that are explicitly stated in the lexicon), and the bonus is missing of being able to guess,

- by induction from the terminological knowledge represented, or
- by deduction from the terminological-cum-morphological knowledge stored in the linguistic knowledge-base.

As a step toward the exploitance of synergism in polyglot knowledge (the whole is not the same as the sum of the components), I have developed a general schema for the representation of linguistic knowledge of a language-family (the application is to the Semitic family).

Now, let us define the basic data-structure that the architecture is defined upon. The main kind of elements in every component of the architecture is, from the viewpoint of the code, relations: flexible, deeply nested relations, coded as nested parentheses. They embody a tree of properties. From the viewpoint of databases, these relations are recursive: every value in a table can be either

- a terminal value,
- or a list of such values,
- or another relation.

From the viewpoint of knowledge-engineering, these nested relations are basically passive frames (accessed by an external control), but control could exploit them as embodying active data, too: actions firing as soon as
they are read. I have applied such data structures to various systems, as
differently interfaced with control (Nissan 1987a-d; 1985); we developed
also a tool for retrieval from such nested relations – RAFFAELLO – first in
Franz Lisp, and then also in Prolog and in C; see (Nissan 1986; 1987e).

We are going to exploit this syntax of representation, for the deviation
model of the Italian regional statutes, too. However, in what we said thus
far about representation, there is nothing that constitutes a deviation model:
indeed, the latter belongs to the data semantics.

In the Semitic multilingual dictionary, lexical frames are associated with
either terms, roots, or root-families. A dichotomy of lexical meanings is
represented inside the same nested relation, possibly together with an
indication of the conjectured historical process of evolution (semantic change)
between the various meanings, as instantiated in the various languages
considered. It is here, that the deviation model belongs; it is structured
along the aforesaid dichotomy.

As to the semantics of the data structures, in the Semitic multilingual
dictionary in particular, the architecture specializes relations for different
roles: it separates lexical entries from semantic entries, with semantics-to-
lexicon entries bridging between the two kinds. This separation between
the lexicon and semantics, and also between lexical entries and semantics-
to-lexicon entries, befits the complexity of an explicit representation of
morphological and semantic relationships inside the lexicon of a coherent
family of languages. Further kinds of frames are specialized for derivational
patterns, as well as for patterns of semantic change. The latter kind, that is
supplemented sometimes by large onomasiological relations, embodies knowl-
dedge about how terms came to have their denotations changed: for example,
the Semitic root gḥḥ is associated, in different Semitic languages, with the
broad concepts “high”, “front”, and “horse”: from “highness” (as instantiated
in Hebrew and Aramaic), it evolved into “front” (as being the highest part
in the head and in the human body); in a certain Ethiopian language, the
senses “with a high front” (also found in Arabic) and “with a broad front”,
are instantiated in derivatives of the same root. One more step is the
Arabic sense “battlefront” (the forefront of the battle), and, given its
employment in the battlefront, “cavalry-front” (Arabic), and finally “horse”
(South Arabian). Close senses such as “horseman” and “horse-cum-horseman
seen as a unit”, coexist in different contexts in Biblical Hebrew.

Not every purpose needs access to such knowledge. Both a cursory
level, and a detail level are provided in most kinds of frames, in our
architecture, in order to subserve superficial requests of access concisely,
and requests for deep knowledge suitably.
Purely semantic frames, according to our approach, develop a dichotomy of description; for example, the frame of the semantic concept "guava" (not the term), contains sets of chunks (forming a tree or a net) that describe the plant, its natural and domesticated distribution, the fruit, varieties and the respective fruits, parts of the fruit, the respective qualities, etc. Practically, the knowledge-base of most frame-based expert systems would fit in the role of purely semantic frames, in the architecture. In given situations, even pictorial data could fit there.

Other kinds of frames envisaged are various kinds of ethnographic frames, such as a kind of frame incorporating a chronology of a given speaking community. There is also a representation of etymological conjectures, and of diverging opinions in etymology.

3.2. Deviation in a metatheory of linguistic inquiry into Etruscan

I have developed and exposed in (Nissan 1989) elements for a computational epistemology of linguistic research meant to interpret the forgotten Etruscan language of pre-Roman Italy. We try to describe patterns of investigation within the trials performed by a single author, as well as the interplay of researchers and schools of research.

One of the reasons for partial, contradictory successes having oftentimes been claimed by Etruscologists, with no convincing breakthrough around, is that the search-space is enormous, but odds are too favorable for having partial matching satisfied for given lexical instances (or – all the more – substrings of a continuous string with no indication of word separation, as frequent in Etruscan inscriptions). Moreover, this is dangerous ground for induction to reconstruct grammar, and for cultural analogy to hypothesize semantics. The enormity of the search-space stems from the fact a relation is sought to one out of very many languages or language-families (known or also practically undeciphered), with conjectures spacing from phonetics and phonology, through morphology (sometimes, fanciful, unwarranted grammars were invented), to lexicography. Failing to have terms matched, led to loosening requirements, by allowing etymologies to be conjectured according to root resemblance and the admission of far-fetched semantic changes; however, this impairs the ability to ascertain the soundness of results.

The metamodel of interpretation, that is, the epistemological model of models, that I have been trying to develop, considers competing partial interpretations of Etruscan as trying to climb to local optima in a universe of paradigms being constructed by learning. Optimality criteria concern the relative and absolute amount of inscriptions completely interpreted, the
robustness of the constructed paradigm (post-optimality) when trials are made to use it to interpret those inscriptions not tackled yet, the ability to dispense with emendations to inscriptions, and the ability to dispense with morphological conjectures that cannot be proved from the linguistic data available. Two main phases are identified in the process of trying to interpret this dead language. The first phase, according to my metamodel, tries to select, among known languages, a candidate model with respect to which to reconstruct deviation. This is attempted by reduced runs of the second phase. Overall paradigm-shifts occur in the first phase. As to the second phase, it bears some relation to inter-intelligibility, as discussed in the former subsection of the present paper. Once a candidate model for deviation is selected (one language, or a family of phylogenetically or historically closely related languages), the second phase focuses search for relatively small discrepancies. A blackboard architecture is considered suitable for representation, if integrating learning methods. There are some analogies with speech-processing, e.g., because the lack of word-separation motivates a so-called “island”-based strategy of word-hypothesizing. However, while English speech-processing refers to a description of English (a known language), for interpreting Etruscan, instead, learning has to bridge between two components: the description (lexicon and grammar) being constructed as a paradigm of Etruscan, with the current partial paradigm being used to constrain further hypotheses, unless the current paradigm is defeated, beyond a certain threshold of perceived failure; and the description (lexicon and grammar) of a known language, selected by the first phase as a model of deviation. It is crucial that an established set of fixed rules of phonetic/phonologic or morphological correspondence would not be departed from. Concluding, the two phases see, respectively, the selection of a language as a deviation model, as opposed to detailed interpretation trials as carried out confidently with respect to a given deviation model.

3.3. Stated versus reconstructed deviation

The difference between the topics of §§ 3.1 and 3.2, in terms of deviation models, is that

- in the former, deviation is stated in the knowledge-representation (this is also the case we are concerned with, in our approach to the representation of the Italian regional statutes),

- while in the case of the metamodel for Etruscology, deviation has to be reconstructed.

The following sections are all devoted to the legal application we are interested in.
4. AN EXAMPLE FROM THE ITALIAN REGIONAL STATUTES

From the statutes of all Italian regions, we are quoting the norms concerning the number of members in the Giunta regionale, that is, the Regional Executive Council; less systematically, we are going to mention the appointment procedure of the Vice President of the Executive Council, to the extent it is specified. Page numbers are from Statuti Regionali (1972). The left column identifies the region.

REGIONS WITH AN ORDINARY STATUTE
(REGIONI A STATUTO ORDINARIO)

ABRUZZO

Art. 38. Composizione della Giunta
La Giunta costituita dal Presidente e da dieci componenti. (p. 25)
(Art. 38. Composition of the Regional Executive Council
The Executive Council is constituted of its President and ten members.)
Art. 48. Vice Presidenza della Giunta
La Giunta provvede alla designazione del Vice Presidente con il compito di sostituire il Presidente in caso di assenza o di impedimento. (p. 28)
(Art. 48. The Vice Presidency of the Regional Executive Council
The Executive Council appoints the Vice President, who has the task of replacing the President in case of absence or impediment.)

BASILICATA

Art. 32. (…) La Giunta composta dal Presidente e da sei assessori, tra cui il Vice Presidente. (p. 44)
(…) The Regional Executive Council is composed of its President and six aldermen, including the Vice President.

According to Art. 34, candidate lists for the Executive Council are voted by the Legislative Council by nominal call and absolute majority. Each list includes the names of the candidate President, Vice President, and aldermen with pre-assigned tasks.

CALABRIA

Art. 17. La Giunta (…) composta dal Presidente (…) e da un numero di assessori non inferiori (sic) ad otto e non superiore (sic) a dodici. (…) Il Presidente della Giunta designa l’assessore Vice Presidente che lo sostituisce in caso di assenza o di impedimento. (p. 62)
(The Regional Executive Council (...) is composed of its President (...) and at least eight and at most twelve aldermen (...)

The President of the Regional Executive Council appoints to the Vice Presidency an alderman who replaces him in case of absence or impediment.)

The norm defining the number of members of the Executive Council, is the same as in Emilia-Romagna.

**CAMPIANIA**

**Art. 32. - Composizione della Giunta**

La Giunta composta dal Presidente e da un numero di Assessori non inferiore ad un decimo e non superiore ad un quinto dei Consiglieri assegnati alla Regione. (p. 84)

(Art. 32. - Composition of the Regional Executive Council

The Executive Council is composed of its President and a number of Aldermen no less than one tenth and not exceeding one fifth of the Legislative Councillors assigned to the Region.)

The latter criterion is similar to the one adopted in Veneto, and is the same as in Toscana. As to the Vice Presidency, according to Art. 33 (p. 85) - Attività e funzionamento della Giunta (Activities and Workings of the Regional Executive Council) - the Vice President is elected by the Executive Council during its first meeting after its composition, on proposal of its President.

**EMILIA-ROMAGNA**

**Art. 25. La Giunta composta dal Presidente e da un numero di assessori non inferiore a otto e non superiore a dodici. (...) (p. 102)**

(The Regional Executive Council is composed of its President and at least eight and at most twelve aldermen. (...))

The norm defining the number of members of the Executive Council, is the same as in Calabria.

**LAZIO**

**Art. 20. La Giunta composta dal Presidente (...) e da un numero di assessori non superiore a 12 (...)**

Il Presidente della Giunta provvede alla designazione dell'assessore vice presidente (...) (p. 121)

(The Regional Executive Council is composed of its President (...) and no more than twelve aldermen (...)

The President of the Regional Executive Council appoints an alderman as vice president (...)
The norm defining the number of members of the Executive Council, is the same as in Piemonte.

Liguria

Art. 37. Composizione della Giunta
La Giunta regionale composta dal Presidente della Giunta e da non più di nove componenti fra cui il Vice Presidente. (p. 142)

(Art. 37. Composition of the Regional Executive Council
The Regional Executive Council is composed of the President of the Executive Council itself, and no more than nine members, including the Vice President.)

According to Art. 38, the Vice President is elected (instead of appointed by the President).

Lombardia

Art. 22. La Giunta composta dal Presidente e da non più di dieci assessori, fra cui il Vice Presidente. (p. 158)

(The Regional Executive Council is composed of its President and no more than sixteen aldermen, including the Vice President.)

According to Art. 23, lists for the Executive Council, to be voted by the Legislative Council, have to indicate the name of the candidate for President, Vice President, and the other members of the Executive Council, with pre-assigned tasks.

Marche

Art. 24. La Giunta regionale composta dal Presidente e da non più di otto Assessori compreso il vice Presidente. (p. 178)

(The Regional Executive Council is composed of its President and no more than eight Aldermen, including the vice President.)

According to Art. 24, the Regional Executive Council votes motions that propose the candidacies for President, Vice President, and the aldermen.

Molise

Art. 20. Composizione della Giunta
La Giunta regionale composta dal Presidente (...) e da non più di otto assessori (…)  
Il Presidente della Giunta designa l’assessore che deve sostituirlo in caso di assenza o impedimento. (p. 196)

(Art. 20. Composition of the Regional Executive Council
The Regional Executive Council is composed of its President (...) and no more than eight aldermen (...)
The President of the Executive Council appoints the alderman who has to replace him in case of absence or impediment.)
The text does not terms that alderman explicitly a Vice President.

PIEMONTE

Art. 31. La Giunta e il suo Presidente
(...)
La Giunta composta dal Presidente e dagli Assessori in numero non superiore a dodici. (p. 217)
(Art. 31. The Regional Executive Council and Its President
(...)
The Executive Council is composed of its President and the Aldermen, whose number should not exceed twelve.)
The norm defining the number of members of the Executive Council, is the same as in Lazio.

PUGLIA

Art. 39. La Giunta regionale composta dal Presidente e da dodici Assessori, fra cui il vice Presidente con funzioni vicarie. (p. 239)
(The Regional Executive Council is composed of its President and twelve Aldermen, including the vice President, who holds vicarious functions.)
According to Art. 40, the President, the vice President and the Aldermen (with pre-assigned sectors of competence) are elected according to a pre-defined global list.

TOSCANA

Art. 36. Numero dei componenti la Giunta
La Giunta composta dal Presidente e da un numero di componenti non inferiore a un decimo e non superiore a un quinto dei Consiglieri assegnati alla Regione. (p. 257)
(Art. 36. Number of Members of the Regional Executive Council
The Executive Council is composed of its President and a number of members no less than one tenth and not exceeding one fifth of the Legislative Councillors assigned to the Region.)
The criterion is the same as in Campania.

UMBRIA

Art. 46. La Giunta composta dal Presidente e da otto membri. (p. 277)
(The Regional Executive Council is composed of its President and eight members.)

According to Art. 48, the candidate Vice President has to be specified in the lists for the Executive Council to be voted by the Legislative Council.

**VENETO**

*Art. 25. La Giunta composta dal Presidente e da un numero di membri non superiore a un quinto dei consiglieri assegnati alla Regione (…) (p. 293)*

(The Regional Executive Council is composed of its President and a number of members not exceeding one fifth of the legislative councillors assigned to the Region (…)*)

The criterion is similar to the one adopted in Campania and Toscana.

**REGIONS WITH A SPECIAL STATUTE**

*(REGIONI A STATUTO SPECIALE)*

**FRIULI-VENEZIA GIULIA**

*Art. 34. Con legge regionale stabilito il numero e sono determinate le attribuzioni degli assessori (…) (p. 315)*

(A regional law establishes the number and determines the tasks of the aldermen (…)*)

**SARDEGNA**

*Art. 37. I componenti della Giunta regionale, preposti ai singoli rami dell’amministrazione, sono nominati dal Consiglio, su proposta del Presidente della Giunta (…) (p. 329)*

(The members of the Regional Executive Council, presiding over the various branches of the administration, are appointed by the Regional Legislative Council, on proposal of the President of the Executive Council (…)*)

**SICILIA**

*Art. 9. Il Presidente regionale e gli Assessori sono eletti dall’Assemblea regionale (…)*

*La Giunta regionale composta dal Presidente regionale e dagli Assessori.*

(p. 336)

(The regional President and Aldermen are elected by the regional Assembly (…)*)
The regional Executive Council is composed of the regional President and Aldermen.

The regional President, in Sicily, is the President of the Regional Executive Council. As to the Vice President, Art. 10 prescribes that:

*Il Presidente regionale in caso di sua assenza od impedimento è sostituito dall’Assessore da lui designato.*

(The regional President, in case of absence or impediment of his, is replaced by one Alderman appointed by him.)

**Trentino-Alto Adige**

Art. 30 (revised in 1971):

*La Giunta regionale è composta del presidente, di due vice presidenti e di assessori effettivi e supplenti.*

*Il presidente, i vice presidenti e gli assessori sono eletti dal consiglio regionale nel suo seno a scrutinio segreto ed a maggioranza assoluta.*

*La composizione della giunta regionale deve adeguarsi alla consistenza dei gruppi linguistici quali sono rappresentati nel consiglio della regione. I vice presidenti appartengono uno al gruppo linguistico italiano e l’altro al gruppo linguistico tedesco.*

*Il presidente sceglie il vice presidente chiamato a sostituirlo in caso di assenza o impedimento.*

*Gli assessori supplenti sono chiamati a sostituire gli effettivi nelle rispettive attribuzioni, tenendo conto del gruppo linguistico al quale appartengono i sostituti.* (p. 355)

(The regional executive council is composed of the president, two vice presidents and actual and substituting aldermen.

The president, the vice presidents and the aldermen are elected by the regional legislative council, from within itself, by secret voting and by absolute majority.

The composition of the regional executive council has to adapt to the consistence of the linguistic groups as represented in the legislative council of the region. The vice presidents belong one to the Italian linguistic group, and the other one to the German linguistic group.

The president chooses the vice president who has to replace him in case of absence or impediment.

The substituting aldermen are appointed in replacement of the actual aldermen in their respective tasks, by holding in due consideration the linguistic group that the replaced ones belong to.)

Trentino-Alto Adige has two autonomous provinces, Trento and
Bolzano (Bozen). The two «linguistic» groups referred to, actually are ethnic groups. The German group is the majority in the province of Bolzano. One third, small group in the region is represented by the few Ladino communities. The statute and the administrative system, and political life in the region are strongly polarized between the two major ethnic groups, unlike the situation in Valle d’Aosta, where bilingualism (Italian and French, with the exception of the juridic system, which is based on Italian) is not accompanied by demographic polarization. This way, Valle d’Aosta has no use for two Vice Presidents.

Valle d’Aosta

Art. 32. Il Presidente della Giunta Regionale, la Giunta e gli assessori che la compongono sono organi esecutivi della Regione.

Art. 33. Il Presidente della Giunta regionale è eletto dal Consiglio fra i suoi componenti, (…)

Gli assessori preposti ai singoli rami dell’Amministrazione sono nominati dal Consiglio su proposta del Presidente della Giunta. (p. 374)

(Art. 32. The President of the Regional Executive Council, the Executive Council itself, and the aldermen who compose it, are executive organs of the Region.

Art. 33. The President of the Regional Executive Council is elected by the Legislative Council from within its own members, (…) The regional aldermen presiding over the various branches of the Administration are appointed by the Legislative Council on proposal of the President of the Executive Council.)

5. Features of the representation adopted

The architecture of our representation is assumed to consist of
– a low-level representation, that is logic-based, and interpreted by Prolog;
– an intermediate representation, that is not suitable for interfacing the user, but is suitable for knowledge-engineering and acquisition. This representation is based on nested relations (or passive frames), of the kind mentioned in Sec. 3.1;
– the syntax of the user-interface.

In this paper, we concern ourselves with the intermediate, frame-based representation. Even though it is flexible (as we are going to see), it is more specialized than the low-level, logic representation, and it assumes more
structure on the part of the knowledge represented. Both the information
stored, and the tools meant to manipulate the intermediate representation,
find support in the low-level, logic representation.

In the intermediate representation, knowledge is structured by attributes,
and under this respect, a tool for retrieval is required. On the other hand,
terminal values can be either numbers, strings, lists of either, or textual
fragments (phrases or simple sentences). A parser is needed for analyzing
these textual data. The parser needs to be logic-based, so it interfaces the
low-level, logic representation. On the other hand, it could be employed
also to handle interaction with users.

A parser answering these needs is being developed in the framework of
another project I have been directing in 1989, ALIBI 2. The tool includes
both a simple Chomskian grammar for syntax, and a deep case semantic
representation. The parser still needs many refinements, but it could be
replaced by a commercial logic based parser.

As to the tool for retrieval from the intermediate representation in nested
relations, it is subserved by the RAFFELLO system; the most suitable version
for this application seems to be the Prolog version implemented in 1989 by
my student Jihad al Sana’ as it allows insertion in a logic-based representa-
tion. However, we have also one C version (developed in 1989 by my
student Ziad Ashkar), and two kinds of implementation in Franz Lisp
(developed by myself, sometimes with the help of different groups of stu-

The data structure, in the intermediate representation, is a recursive list
of lists, that embodies a tree of properties. (Besides, whatever follows a
semicolon on the same line is considered to be comment, and is ignored
during retrieval.) Each property is identified by an attribute, that usually
is explicitly named in upper-case, but that can be implicit, in certain contexts
that we are going to specify. Each property is delimited by the parenthesis
enclosing it. This parenthesis includes the name of the attribute first,
followed either by one or more terminal values, or by one or more inner
properties. These nested properties can be all different (every kind occurs
once), so they are explicitly identified by the upper-case name of their
respective attribute; Example 1 follows:

(ATTRIBUTE_1
 (ATTRIBUTE_2 (value_a value_b))
 (ATTRIBUTE_3 (value_c))
 (ATTRIBUTE_4
 (ATTRIBUTE_5 (value_d value_e))
 (ATTRIBUTE_6
Otherwise, the nested properties can constitute a sequence of one or more instances of one or more attributes. If several instances of just one attribute occur in the same sequence, then it is superfluous to state the name of the attribute explicitly; each property will be an *unnamed chunk*, that can contain either terminal values, or further nested properties. For example, in the following piece of code, two unnamed chunks are nested inside the property named after `ATTRIBUTE_1`, and each has the same structure, in the sense that all or some of the attribute nested inside one chunk, occur (with possibly different values) in the following chunk (*Example 2*):

```plaintext
(ATTRIBUTE_1
 (ATTRIBUTE_2 (value_a value_b))
 (ATTRIBUTE_3 (value_c))
 (ATTRIBUTE_4.......)
 (ATTRIBUTE_5.......)
)

((ATTRIBUTE_2 (value_d))
 (ATTRIBUTE_3 (value_e))
 (ATTRIBUTE_5.......))
```

One or more of the attributes have their value providing the identification key of the repeated chunk. For example, let `ATTRIBUTE_3` in the example be the key, but a key could be even multiple, and involve several (or even all) of the attributes occurring in a given chunk.

Unnamed chunks introduce, in our syntax, *multisets* (as opposed to *simple sets*: sets of elements that can occur just once). Indeed, unnamed chunks allow several instances of the same (implicit) attribute to be nested side by side. However, the limitation is that several occurrences are allowed for just one attribute. The most general case of *multiset* is allowed by the following syntactic feature; indeed, we are interested in admitting repeatable chunk-patterns even as instances of one out of a set of attributes; in such cases, it is obvious that the name of the relevant attribute notwithstanding
the repetition must be stated explicitly (except at most one of the attributes), as in the following *Example 3*:

```
(ATTRIBUTE_1
 (ATTRIBUTE_2 (ATTRIBUTE_3 (value_a))
 (ATTRIBUTE_4 (value_c))
 (ATTRIBUTE_5......)
 (ATTRIBUTE_6......))
)
(ATTRIBUTE_2 (ATTRIBUTE_3 (value_d))
 (ATTRIBUTE_4 (value_e))
 (ATTRIBUTE_6......)
)
(ATTRIBUTE_7 (ATTRIBUTE_8 (value_f))
 (ATTRIBUTE_9 (value_g))
)
(ATTRIBUTE_2 (ATTRIBUTE_3 (value_h))
 (ATTRIBUTE_5......)
)
(ATTRIBUTE_7 (ATTRIBUTE_8 (value_i))
 (ATTRIBUTE_9 (value_j))
))
```

We could have omitted the name of either ATTRIBUTE_2 or ATTRIBUTE_7, because the occurrence of two open parenthesis characters, successively with no other characters between (excepts blanks, newlines, or comments following a semicolon), at the level nested immediately inside ATTRIBUTE_1, makes sure the first parenthesis opens an unnamed chunk, to be associated with the one implicit attribute allowed.

The *Prolog* version of *RAFFAELLO* converts the syntax of our nested relations (originally developed in a *Lisp* environment) into nested lists as in a syntax that *Prolog* can handle; then, retrieval from the nested relation is performed, according to a database schema (*metarepresentation*) that specifies legal nesting of properties.

A query addressing the intermediate representation is a predicate that takes a list as argument; the list is the *retrieval path* through the nested levels of the frame: the path consists of the name of a frame, followed by a sequence of names of nested attributes, from the outermost, to the innermost interested. Incomplete but unambiguous paths are also allowed. As to traversing an unnamed chunk, instead of naming the attribute (that in this case is implicit), one has to state the key value (that is identified as such because unlike names of attributes, it includes at least one lower-case letter). As to the case when repeated chunks of different attributes occur that is,
in the generalized multiset case the path has to include the name of the attribute, followed by the key value that identifies the particular chunk.

The retrieval tool consults the metarepresentation, and is able to analyze its syntax: we have defined a language, named CuProS, short for Customization Production System, as it is rule-based, and customizes the retrieval tool for the particular application. Each rule describes legal nesting in a given attribute, that is named in the left side; implicit attributes (those with unnamed chunks) have a conventional name, in use only inside the metarepresentation, and written (at least partly) in lower case.

The nested relation itself is at the object-level of representation, as opposed to the meta-level representation (or metarepresentation). Each kind of nested relation may have several instances at the object level, and one metarepresentation. Example 4 is a piece of metarepresentation, concerning the object-level code of Example 1:

```
(ATRIBUTE_1 (ATTRIBUTE_2
ATTRIBUTE_3
ATTRIBUTE_4
ATTRIBUTE_8
))
(ATRIBUTE_4 (ATTRIBUTE_5
ATTRIBUTE_6
))
(ATRIBUTE_6 (ATTRIBUTE_7
......
))
```

As to Example 2, its metarepresentation in the following Example 5:

```
(ATRIBUTE_1 (n: chunk_under_Attribute_1))
\langle chunk_under_Attribute_1 (ATTRIBUTE_2
i: ATTRIBUTE_3
ATTRIBUTE_4
ATTRIBUTE_5
))
```

In the latter example, we see that the right part of rules can be interspersed with labels (ending by a colon); they are meant to convey information concerning the following attribute, or – as the case is of other labels, such as x: (that corresponds to an exclusive or of substructures of frame) – of a sub-unit of the right part of the rule, listing one or more attributes.

The label n: indicates that the following attribute can have repeated
chunks; in the example, that attribute has its name containing lower-case letters (only one letter happens to be capitalized), thus the repeatable chunk is an unnamed chunk, at the object level.

The label \( i2 \) indicates that the attribute that follows it, is the key attribute of the chunk-pattern. Instead, according to the code in the following Example 6, the key includes two attributes, preceded by the label \( i2 \):

\[
\begin{align*}
(\text{ATTRIBUTE}_1 & \ (n: \text{chunk}_\text{under}_\text{Attribute}_1)) \\
(\text{chunk}_\text{under}_\text{Attribute}_1 & \ (i2: \text{ATTRIBUTE}_2) \\
i2: \text{ATTRIBUTE}_3 & \\
\text{ATTRIBUTE}_4 & \\
\text{ATTRIBUTE}_5 & \\
) \\
\end{align*}
\]

The code in the following Example 7, is the metarepresentation of the object-level code of Example 3:

\[
\begin{align*}
(\text{ATTRIBUTE}_1 & \ (n: \text{ATTRIBUTE}_2) \\
n: \text{ATTRIBUTE}_7 & \\
) \\
(\text{ATTRIBUTE}_2 & \ (\text{ATTRIBUTE}_3) \\
\text{ATTRIBUTE}_4 & \\
\text{ATTRIBUTE}_5 & \\
\text{ATTRIBUTE}_6 & \\
) \\
(\text{ATTRIBUTE}_7 & \ (\text{ATTRIBUTE}_8) \\
\text{ATTRIBUTE}_9 & \\
) \\
\end{align*}
\]

6. EXAMPLES OF FRAMES AND THEIR METAREPRESENTATION

The following Example 8 is the nested relation of the concept \( \text{Giunta}_\text{regionale} \), that is, «Regional Executive Council». In the code shown, we are going to provide in detail only the representation of knowledge on how many members (regional aldermen, \( \text{assessori alla Regione} \)) belong to the Executive Councils of the various Regions, according to the excerpts from the regional statutes, seen in Sec. 4:

\[
\begin{align*}
\text{assign_frame} \ & \text{Giunta}_\text{regionale} \\
(\text{DEFINITION} & \ ((\text{Every region has a Giunta}_\text{regionale}) \\
\text{The Giunta}_\text{regionale} \ & \text{is the Executive organ of the Region}) \\
\text{The Giunta}_\text{regionale} \ & \text{has one Presidente_della_Giunta and many an} \\
\text{Assessore_alla_Regione} & \ (\text{SEE ATTRIBUTE ALDERMEN}) \\
) \\
\end{align*}
\]
(The Giunta_regionale usually has one Vice_Presidente_della_Giunta)
)
(APPOINTMENT_MODALITIES .................)
(COMPETENCES .................)
(ALDERMEN
(HOW_MANY
((IF (Region is Abruzzo))
(VALUE_IS (10))
)
((IF (Region is Basilicata))
(VALUE_IS (6))
)
((IF (Region is Calabria
or Emilia-Romagna
))
(VALUE_IS
(AND
(>= 8); the same as: (at least 8)
(<= 12); the same as: (at most 12)
))
)
((IF (Region is Campania or Toscana))
(VALUE_IS
(AND
(>= \#X / 10)
(<= \#X / 5)
(DEFINE
(\#X is how many a
Consigliere_regionale
the Region has
))))
)
((IF (Region is Lazio or Piemonte))
(VALUE_IS (at most 12))
)
((IF (Region is Liguria))
(VALUE_IS (at most 9))
)
((IF (Region is Lombardia))
(VALUE_IS (at most 16))
)
((IF (Region is Marche or Molise))
(VALUE_IS (at most 8))
)
(((IF (Region is Puglia))
(VALUE_IS (12))
)

(((IF (Region is Umbria))
(VALUE_IS (8))
)

(((IF (Region is Veneto))
(VALUE_IS (at most one fifth of
how many a Consigliere_regionale
the Region has)))
..........)
); end of HOW_MANY
..........)
); end of ALDERMEN
..........)
); end of frame.

The format adopted for the chunk of Veneto is more readable than the format seen for Campania and Toscana.

A relatively more concise representation, that, however, would require a separate formulation of the criterion of interpretation, is the following, in Example 9:

(ALDERMEN
(HOW_MANY
((CRITERION (fixed number))
(SCHEMA_OF_ATTRIBUTES ((Region) (Value)))
(TYPE_OF_ATTRIBUTES ((strings) (number)))
(RELATION
((Abruzzo) (10))
((Basilicata) (6))

((Puglia) (12))
((Umbria) (8))
))

((CRITERION (at most))
(SCHEMA_OF_ATTRIBUTES ((Region) (at most)))
(TYPE_OF_ATTRIBUTES ((strings) (number)))
(RELATION
((Lazio Piemonte) (12))
((Liguria) (9))
In the case of the criterion *constantly delimited range*, only one instance – the chunk comprising Calabria and Emilia-Romagna – occurs under RELATION; thus, having to define three more properties (CRITERION, SCHEMA_OF_ATTRIBUTES, and TYPE_OF_ATTRIBUTES) according to the new format of Example 9, yields, for the instance considered, a representation that is even more prolix than in Example 8. Then, for the situation when the number of regional aldermen is constantly and explicitly delimited by a lower and an upper bound, the chunk identified by the key property ((CRITERION constantly delimited range)), and repeated hereby for the reader’s convenience (Example 9):

```
((CRITERION (constantly delimited range))
 (SCHEMA_OF_ATTRIBUTES ((Region) (at least) (at most)))
 (TYPE_OF_ATTRIBUTES ((strings) (number) (number)))
 (RELATION ((Calabria Emilia-Romagna) (8) (12)))
 )
```
could be usefully replaced by (Example 10):

```
(\{(\text{Region} \text{ is Calabria or Emilia-Romagna})
   \land \text{(at least 8) (at most 12))}
```

in the framework of the code of Example 9, but according to the format of Example 8.

According to the syntax of the metarepresentation language we defined, CuProS, such flexibility in the format of the frame portion under HOW_MANY, is expressed, in a global metarepresentation file, as follows (Example 11):

```
(executive_organism; It is the kind of frame that
(DEFINITION; Giunta_regionale belongs to:
  APPOINTMENT_MODALITIES; the attributes may appear at
  COMPETENCES; the upper level of nesting.
  \(x\): ALDERMEN; \(x\): stands for «either... or...»
  \(x\): MINISTERS; and delimits one or more attributes.
))
(ALDERMEN (HOW_MANY........))

(HOW_MANY (m: if-value_chunk
n: criterion-dependent_chunk))
```

```
(if-value_chunk (i: IF; key.
VALUE_IS))
```

```
(criterion-dependent_chunk
(i: CRITERION; key.
DEFINE
par: SCHEMA_OF_ATTRIBUTES (applies_to: RELATION)
par: TYPE_OF_ATTRIBUTES (applies_to: SCHEMA_OF_ATTRIBUTES)
multi-par: RELATION
))
```

The object-level representation under HOW_MANY could be augmented, with a bibliographic reference to the particular article(s) of the particular regional statute, where the information represented is stated. This could be done by inserting a REFERENCE attribute inside if-value chunks and criterion dependent chunks. Then, the metarepresentation rule describing if value chunks should be augmented as follows:

```
(if-value_chunk (i: IF
VALUE_IS
REFERENCE))
```
The \textit{par:} labels, in the last rule of \textit{Example 11}, state that the ordered set of elements given, in the object level representation, as value of the attributes \texttt{SCHEMA\_OF\_ATTRIBUTES} and \texttt{TYPE\_OF\_ATTRIBUTES}, have to be considered as parallel. Indeed, as we have seen, for instance, in \textit{Example 9}, we have the correspondence:

\begin{verbatim}
(SCHEMA\_OF\_ATTRIBUTES ((Region) (at least) (at most)))
(TYPE\_OF\_ATTRIBUTES ((strings) (number) (number)))
\end{verbatim}

where the second property defines the type of the elements listed under the first property, that, in turn, is the schema of a tabular relation, stated under \texttt{RELATION}.

Instead, the label multi-par: that, in the metarepresentation formulated in the last rule of \textit{Example 11}, precedes the attribute \texttt{RELATION}, states that, in the object level representation, the property \texttt{RELATION} consists of the name of the attribute \texttt{RELATION} followed by a sequence of rows in a table, that is, followed by a list of elements, each being parallel to those properties, whose attribute name in the same metarepresentation rule is preceded by the label \texttt{par:}

To specify some semantics for the attributes of the right part of the rule, the optional, parenthesized statement – \texttt{applies to:} – that accompanies the \texttt{par:}-labelled attributes, stipulates, in the last rule of \textit{Example 11}, that the values of \texttt{TYPE\_OF\_ATTRIBUTES} applies to the respective elements of \texttt{SCHEMA\_OF\_ATTRIBUTES}, that, in turn, apply to the respective elements of the elements (rows) of \texttt{RELATION}: the inner level (elements of elements, instead of elements), is conventionally due to \texttt{multi-par:} having been used as label before \texttt{RELATION}, instead of \texttt{par:}

The intricacy of the \texttt{par:} and \texttt{multi-par:} syntax is due to the fact that they define the metarepresentation of properties that, in turn, are a local metarepresentation inserted in object level representation code (indeed, \texttt{SCHEMA\_OF\_ATTRIBUTES} defines a syntactic schema for attributes included in the same chunk of object-level representation, while \texttt{TYPE\_OF\_ATTRIBUTES} defines some \textit{data semantics}, corresponding to the local syntactic metarepresentation provided by \texttt{SCHEMA\_OF\_ATTRIBUTES}). These features of our metarepresentation language are meant to enhance its expressive flexibility, but are not indispensable, if one adopts our representation formalism: indeed, one can adopt the schema of \textit{Example 8}, that is simpler than the schema of \textit{Example 9}.

Some explicit statement, in the \textit{global} metarepresentation, of the \textit{data semantics}, is useful (but not indispensable, if we chose to embed, implicitly,
knowledge on the data types in control modules of the expert system exploiting the nested relations. Please consider the explicit data-semantics metarepresentation in Example 12:

(IF (text))
(CRITERION (text))

(VALUE_IS (x: number; x: stands for
x: text; «exclusive or».
))

(SCHEMA_OF_ATTRIBUTES (local_metarepresentation))
(TYPE_OF_ATTRIBUTES (data_types))

The last line states that the attribute TYPE_OF_ATTRIBUTES can have values that have to be interpreted as being data types. Indeed, TYPE_OF_ATTRIBUTES belongs in a local semantic metarepresentation.

Deeper semantics – not just the data types – of attributes could be stated separately, but, if the names of the attributes are meaningful, one can conceive some module (that we don’t have), inside the natural-language parser, that could handle the meaning of such attribute names such as APPOINTMENT_MODALITIES.

As to the line (SCHEMA_OF_ATTRIBUTES (local_metarepresentation)), it stipulates that the property SCHEMA_OF_ATTRIBUTES is a piece of code that appears inside a nested relation at the object-level of representation, yet its value has to be interpreted as a (local) metarepresentation.

The first rule of Example 11 describes the upper level of frames whose kind is executive_organism. We assume that the nested relation Giun-\textit{ta\_regionale} of Example 8 belongs to this kind of frame. However, the retrieval manager needs to be told about this assumption. Then, we provide a nested relation whose role is to (meta)describe the association between available nested relations of the object-level of representation, and kinds that have a global metarepresentation associated (each in a file on its own, or all together in one file). See the following Example 13:

(assign_frame list_of_frames
((IDENTITY (list_of_frames))
(AVAILABLE FRAMES; Let the list be:
((OF_KIND (executive_organism)
(ARE
(Giunta\_regionale; Regional Executive Council.
Consiglio\_dei\_Ministri; National Government.
)))
The following Example 14 is the metarepresentation of the previous nested relation:

```
(ball_of_flowers (list_of_objects)
  (atom (IDENTITY AVAILABLE_FRAMES))
)
```

Now, let us consider the frame of *Vice_Presidente_della_Giunta*. Some properties are shared by a Vice President of a Regional Executive Council (*vicepresidente della Giunta regionale*), and a Vice President of a Municipal
Executive Council (vicepresidente della Giunta comunale or municipale), so let us define one nested relation to describe both roles, with all differences (the differentia specifica with respect to the common genus) described in two acceptation-chunks inside the property ACCEPTATIONS. The fact that Trentino-Alto Adige has two regional vice presidents (one for each major «linguistic» or ethnic group) provides us with the opportunity of exemplify default values (here, one vice president) as opposed to exceptions. See the following Example 15:

(assign_frame Vice Presidente_della_Giunta
  (DEFINITION......)
  (COMPETENCES......)
  (ACCEPTATIONS
    ((OF (Giunta_comunale)); key.
      ......
    )
    ((OF (Giunta regionale)); key.
      (HOW_MANY
        (DEFAULT (1))
        (EXCEPTIONS
          (IF (Region is Trentino—Alto_Adige))
          (THEN
            (REPLACING_VALUE (2))
            (PRESCRIPTION
              (The two must be one ethnic_Ionian
               along with one ethnic_German
               )))
          (APPOINTMENT_MODALITIES......)
          (COMPETENCES......)
        )
      )
    )
  )
)

Let us add a property that, from the viewpoint of legal applications, is rather unimportant, but that is more interesting from the viewpoint of orthography, as involved, for example, in corpus linguistics: the various Italian regional statutes adopt different spellings, when referring to the Vice President: Vice Presidente, vice Presidente, Vice Presidente, and vice presidente. Besides, informally (such as in newspapers), the most likely spelling is vicepresidente. One out of the many representational choices could be as in the following Example 16:

(SPELLING
  (FORMAL
    (SCHEMA_OF_ATTRIBUTES

Lexical differences (as opposed to orthographic differences) would be stated more usefully stated. In Sicily, the Presidente regionale corresponds to the Presidente della Giunta regionale of other regions. In Valle d’Aosta, the Regional Legislative Council is termed Consiglio della Valle (literally: the Council of the Valley), instead of Consiglio regionale, which is the compound is use for other regions. The natural-language processing component could exploit idiosyncratic knowledge on the pragmatic use of morphology, when in connection with the name of a particular region. The knowledge about:

- the official name of Valle d’Aosta being Valle d’Aosta, while the usual informal name is Val d’Aosta;
- the definite article being needed before both forms (la Valle d’Aosta, la Val d’Aosta), which implies that, because of the composition of prepositions with the definite article according to the general grammar of Italian, the genitive is della Valle d’Aosta versus della Val d’Aosta;
- when co-occurring with in, the definite article is omitted in one case, while being kept (and composed) in the other: nella Valle d’Aosta, versus in Val d’Aosta.

Such idiosyncratic knowledge could be stated in a specific property on linguistic use, to be included in the nested relation of the Region: Valle_d_Aosta. The property would be accessed by the parser. Particular conventions about transcription should obviate the inability of Prolog to handle properly certain spellings, if used as being the names of atoms: there are ways to handle the spelling of text properly.
Thus far, we dealt with the intermediate representation. Instead, the lowest level of representation is intended to be logic-based, with semantic predicates such as the following, simple ones:

\[ \text{has}(X,Y):= \text{governs}(X,Y). \]
\[ \text{governs}(\text{giunta_regionale}, \text{region}). \]
\[ \text{governs}(\text{actor}, \text{object}): \ldots \ldots \]

There could be knowledge about people in office at the giunta_regionale at a given time:

\[ \text{in\_office}(\text{giunta\_regionale}, \]
\[ [ [ \text{president}, \text{tom} ], \]
\[ [ \text{aldermen}, [ \text{dick}, \text{harry}, \text{etc} ] ] \]
\[ ] \]
\[ \text{region\_is}(\text{lombardia}), \]
\[ \text{period\_is}(1972, \text{to}, 1976). \]

To handle semantics more generally, a predicate provides a general schema for action verbs. The order of arguments is conceptually (though not from the viewpoint of efficient pattern-matching) irrelevant, as each argument is, conventionally, a list whose first element identifies the kind of deep case. The number of arguments that actually appear is free. The schema is:

\[ \text{verb}([ \text{action}, \text{govern } ], \]
\[ [ \text{actor}, \text{Actor } ], \]
\[ [ \text{object}, \text{Object } ], \]
\[ [ \text{location}, \text{Location } ], \]
\[ [ \text{time}, \text{Time } ], \]
\[ \ldots \ldots \). \]

7. Conclusions

We have exposed a knowledge-representation model, adapted for legal applications. The architecture comprises an intermediate representation (under the user interface), and a logic-based low level. In turn, the intermediate representation includes

1) an object level (based on deeply nested relations), and
2) a rule-based meta-level, that is structured as
   - a separate, global description of attributes and nesting, formulated in the CuProS metarepresentation language,
   - but possibly also as a local schema of attributes, of interest in a particular frame.
Of this metarepresentation, we have discussed especially ways to stipulate the syntax of attributes, but we have also dealt with semantics, especially data types.

The particular application discussed, is the Italian regional statutes. We have proposed a unified representation of the semantics of these more or less parallel corpora, and have discussed other forms of deviation models, as devised in other projects developed by the author.

REFERENCES


